Learning to Classify Objects from Real Visual Images: a Biologically Inspired Approach

Manuel A. Sánchez-Montañés, Luis F. Lago-Fernández, Fernando J. Corbacho

Abstract— **Biological systems have a remarkable ability to learn from the world. As engineers, we try to build autonomous agents that interact and adapt in open, partially unpredictable environments; our approach consists in looking into biology trying to understand and generalize some of the crucial mechanisms that are neccesary for learning and adaptation. This paper describes an artificial neural network that learns to classify different visual stimuli using both supervised and unsupervised mechanisms. The system consists of a segmentation network that separates the different objects in a visual scene; a network that learns internal representations of the visual stimuli; and a network that learns to classify the acquired internal representations. The structure of the system and all its mechanisms are inspired by biological findings. We tested the system using real grey-level images from a camera. The result is a system that is robust against noise and learns to classify objects independently of their position with respect to the camera.**

Keywords— **Unsupervised/Supervised Learning, Neural networks, Computer Vision, Inspiration from Biology.**

I. Introduction

Despite all the scientific and engineering effort in the last decades, biological systems remain orders of magnitude more robust, adaptive and flexible than artificial ones [1]. In this paper we introduce an artificial learning system for classification of visual images that incorporates several of the principles of organization that have proved essential for the success of autonomous biological agents. It incorporates a visual system capable of robust object segmentation and recognition based on very adaptive neural principles; it includes synchronization in cortical structures to achieve segmentation of visual scenes and unsupervised learning of internal representations of the different objects views. These representations are then associated to their corresponding class through a supervised learning mechanism.

II. Overview

On the first stage of the process a gray level image of 320×240 pixels is captured by the camera. The image is filtered and processed in order to reduce it to a matrix of ones and zeros, which will be used as input for the next steps. In figure 1 we show a general schema of the whole system. After the initial pre-processing, segmentation is performed using a low resolution version of the image. Once the objects have been separated they are used to filter the original image. This filtered image, that keeps the initial resolution, will be used as input to the subsequent layers. The last step before it can be recognized is a normalization that provides position invariant representations of the objects. In the following sections we describe in detail each of the mentioned processing steps.

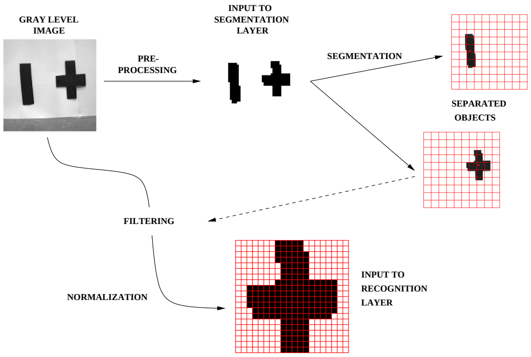

Fig. 1. Schematics of the visual system.

III. Object segmentation

The image is segmented using a network in which the different objects (connected regions) in the visual scene are represented by groups of synchronized neurons. In the last years there has been an increasing insterest on the role synchronization should play in the processing of information by neural systems ([2], [3], [4]). Therefore, using a synchronization-based solution to the problem of visual segmentation we are going a step closer to the robustness of biological systems. The network is based on previous works ([5], [6]). Each pixel in the image is connected as input to a pair of one excitatory (A) and one inhibitory (B) integrate and fire neurons mutually connected. A units are locally coupled to nearest neighbors and connected to a population of inhibitory neurons (C) that send inhibition to all the network (see figure 2). The dynamics of all the units in this network are given by:

$$\frac{dV_i}{dt} = -\gamma_i(V_i - V_i^{rest}) + \sum_{j \in \Gamma(i)} g_{ji}(V_i - V_{ji}) \qquad (1)$$

where γ_i represents the leakage conductance and V_i^{rest} the membrane equilibrium potential of the neuron i. The sum extends to all the neurons connecting the neuron i, g_{ji} being the synaptic conductance and V_{ji}

The authors are with the E. T. S. de Informática, Universidad Autónoma de Madrid, Spain. E-mail: Manuel.Sanchez-Montanes@ii.uam.es .

the synaptic reversal potential. After a presynaptic spike the conductance is assumed to rise instantaneously from 0 to the value g_{ji}^{max}, and to keep this value for a short period of time T_{ji}, according to the equation:

$$g_{ji}(t) = g_{ji}^{max}[\Theta(t - t_j^{last})(1 - \Theta(t - t_j^{last} - T_{ji}))] \quad (2)$$

where $\Theta(x)$ is the Heaviside function and t_j^{last} is the time of the last spike of the presynaptic neuron j. The parameters in the previous equations vary depending on the type of unit and the type of connection, and are tuned so that:

1. Only units connected to an active pixel can fire.
2. When a A unit fires it excites its corresponding B unit, its neighbor A units and the population of C units.
3. When a B unit fires it inhibits its A unit, giving rise to a period of inactivation of unit A.
4. When the C units fire they inhibit the A units, inactivating all of them except those receiving a high synaptic current from their neighbors.

With all these elements, the segmentation mechanism makes all the A neurons belonging to the same object fire synchronously; the competition implemented through the global inhibitor C guarantees that neurons responding to different objects fire at different times.

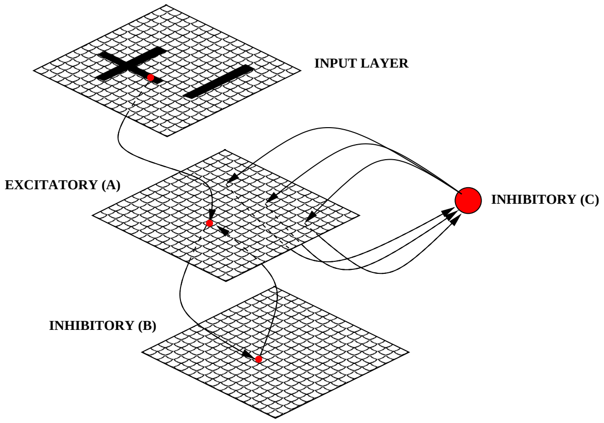

Fig. 2. Structure of the segmentation network. Local connections between E units have been omitted for clarity.

IV. Learning to classify the objects

On the next stage a second network processes the segmented images, previously normalized in order to have position-independent representations of the objects. This network is composed of a first part that acquires internal representations of the objects and a second part that associates classes to these internal representations.

A. Learning internal representations of the objects

The first part of the learning network is inspired by previous models of self-organization in cortex ([7], [8]) that learn to discriminate between the different stimuli in the environment.

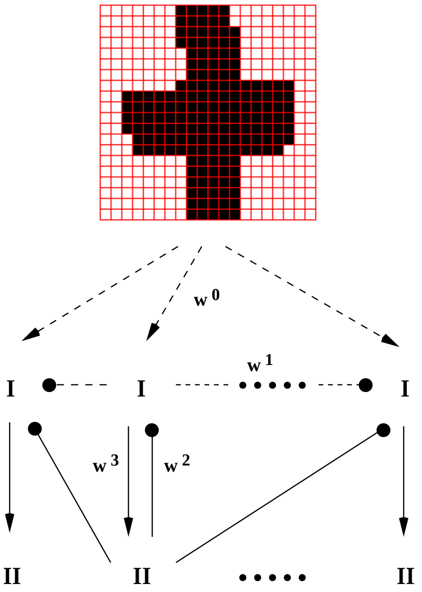

Fig. 3. Structure of the recognition network. Excitatory connections are represented by arrows, inhibitory ones by circles. All dashed connections are plastic.

There are two types of units (output neurons, I, resembling pyramidal neurons in the biological system, and inhibitory interneurons, II), both with firing rate dynamics. They are modelled by the equations:

$$\begin{aligned} V_j^I(t+1) &= \alpha V_j^I(t) + \gamma_0^j \sum_{i=1}^{N_{in}} w_{ij}^0 A[V_i^0(t)] \\ &- \gamma_1 \sum_{i \neq j}^{N_I} w_{ij}^1 A[V_i^I(t)] \\ &- \gamma_2 \sum_{i=1}^{N_{II}} w_{ij}^2 A[V_i^{II}(t)] \qquad (3) \\ V_j^{II}(t+1) &= \alpha V_j^{II}(t) + \gamma_3 w_{jj}^3 A[V_j^I(t)] \qquad (4) \end{aligned}$$

where $V_j(t)$ is the level of excitation of unit j, α is a constant between 0 and 1, $V_i^0(t)$ is the level of activation of unit i in the input layer, and γ and w_{ij} are the gain and the strength of the corresponding connections. $A[X]$ is the output activity of a unit with excitation level X, defined as:

$$A[X] = \Theta(X - V_{Th})X \qquad (5)$$

where $\Theta(x)$ is the heaviside function, and V_{Th} is the firing threshold of the unit. Each I unit receives feedforward connections from the whole input layer (segmented normalized image), lateral connections from the other I units and inhibitory connections from all the II units. Each II unit receives excitation from a I unit (see figure 3). The connections from the input layer to layer I are initially randomly distributed between 0 and 1, making all the I units excitable by any visual stimulus. However, these connections are modifiable by experience, increasing the selectivity and specificity of I units: if at the moment t when the unit j in layer I starts firing ($A[V_j^I(t)]$ becomes greater than 0) its inhibitory input is smaller than a fixed value C, then the synapses it receives from the input layer are subject to change (the unit 'learns'). This learning constraint is biologically plausible as discussed in ([8], [9]). Its role is to focus the learning to a small subset of the I units that initially respond to the stimulus, 'reserving' the other units for learning different stimuli ([8], [9]). In order to make the learning in the network robust against transients, a synapse also needs to be activated during at least a period of time P to be modified. Then,

$$\Delta w_{ij}^0(t) = \alpha \qquad (6)$$

The sign of α is positive when $A[V_i^0(t)] > 0$, and negative if $A[V_i^0(t)] = 0$ (homosynaptic LTP and heterosynaptic LTD in biological terms, [8]). This mechanism makes the response of the unit more specific to the input stimulus. The strengths of these synapses are constrained within the $[-1, 1]$ range. The gain γ_0^j also adapts when the unit is learning:

$$\Delta\gamma_0^j(t) = \beta(V_{Max} - V_j^I(t)) \qquad (7)$$

This mechanism normalizes the excitation a unit receives. Therefore, when a unit has learned its optimal stimulus, the level of excitation it receives when this stimulus is present is near V_{Max}, independently of the number of input units that code the visual stimulus, preventing malfunctioning and network overexcitation. If this mechanism is not present, a unit that learns to fire to a cross would have strong excitatory input, firing also to a small part of it (like the vertical bar). The gain γ_0^j and the connection weights w_{ij}^0 are 'frozen' when the learning dynamics for the I unit j has converged. This makes the network more stable. Finally, lateral inhibitory connections between units I are initially null, being modifiable through experience. They implement competition between these units. Hence, they allow a subset of units that learn to recognize a particular stimulus (and are responding to it) to decrease the activity in the other units. In case $A[V_j^I(t)] > 0$ and the *ith* neuron in layer I learns the stimulus during at least a period of time P the synapse w_{ij}^1 is strengthened.

B. Association of classes to internal representations

We use a third neural network that learns to classify an internal representation of the object. It learns to associate visual stimuli with their corresponding classes. The dynamics of the neurons are the same as in the previous network. In figure 4 we represent schematicaly the structure of this network.

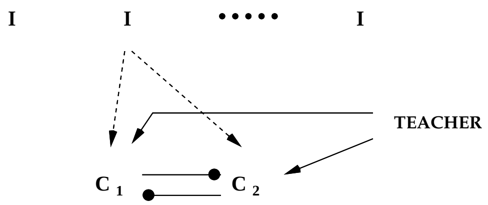

Fig. 4. Structure of the classification network. When the teacher externally activates one of the C_i units in the presence of a neutral stimulus, the connection between this unit and the neuron coding the stimulus in layer I is strengthened.

We consider two different units (C_1, C_2) used by the system to classify the stimulus as 'bad' or 'good' respectively. Implicitly the system classifies the stimulus as neutral in case none of the C_i units is active. There is mutual inhibition between C_1 and C_2, so that the most active C_i stops the activation of the other, hence having the network perform a single classification. C_i units receive connections from layer I. These connections associate a visual stimulus (coded by layer I) with the corresponding object class. Initially, these connections are null, which makes all the visual simuli initially neutral (none of them activates the C_i units). When the system is processing a novel object the external teacher can classify it by activating one of the C_i units. Because the connection from unit i in layer I to C_j is strenghtened if they are simultaneusly active, the network associates the unit in I that represents the visual input from the object with its class. With this mechanism the system learns to classify objects by activating units C_i with no need of the teacher, predicting when the same visual stimulus is present the class the teacher would have given.

V. Results

To test the system we have used grey-level images from a camera. The segmentation network was able to separate appropiately different objects in the same image, and the internal representations learned in the second network showed to be robust against noise: a unit that learns to detect a visual stimulus also detects perturbations of that image in certain range. Because of this just a few units are neccesary to detect an object in any position. The association of the internal representation of the object view with the object class was succesfully implemented: for novel objects the information about their class was given to the system, and the network was able to perform correct classifications on subsequent perturbed and unlabeled presentations. Finally we tested the system on an autonomous robot with a camera as its only sensor. The visual information processed by our model proved to be sufficient for the robot to achieve simple tasks in a controlled environment (e. g. avoiding obstacles and approaching objectives).

VI. Conclusions

In our experiments we considered just two classes but the architecture is perfectly extensible to N classes. We conclude that the properties showed by the system make it useful for real world applications, e. g. artificial vision systems. In addition the parallel structure of the system makes it very suitable for parallel VLSI implementation. Finally we suggest that the study of biological systems can give rise to the development of new architectures that process sensory information and learn optimally from complex environments.

Acknowledgments

We would like to thank the Comunidad de Madrid for a F. P. I. grant to L. F. L.

References

[1] Corbacho, F. and Arbib, M. A., *Towards a Coherence Theory of the Brain and Adaptive Systems*, Proceedings of the First International Conference on Vision, Recognition and Action (ed. Stephen Grossberg). Boston, MA, 1997.

[2] von der Malsburg, C., *The correlation theory of brain function*, Max-Planck Internal Report, 1981. Reprinted in: Domany, E., Van Hemmen, J. L. and Schulten, K. editors, Models of Neural Networks II. Springer-Verlag, 1994.

[3] Gray, C. M., König, P., Engel, A. K. and Singer, W., *Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties*, Nature 338: 334-337, 1989.

[4] Gray, C. M., *Synchronous oscillations in neuronal systems: mechanisms and functions*, J. Computational Neuroscience 1: 11-38, 1994.

[5] Wang, D. and Terman, D., *Image Segmentation Based on Oscillatory Correlation*, Neural Comp. 9, 805-836, 1997.

[6] Lago-Fernández, L. F., Sánchez-Montañés, M. A., Corbacho, F. and López-Buedo, S., *A Biologically Inspired Autonomous Robot that Learns Approach-Avoidance Behaviors*, Proceedings of the Fourth International Conference on Autonomous Agents, ACM Press, New York, 2000.

[7] Sánchez-Montañés, M. A., Corbacho, F. and Sigüenza, J. A., *Development of Directionally Selective Microcircuits in Striate Cortex*, Proceedings of IWANN-99. Lecture Notes in Computer Science 1606. Springer Verlag: Berlin, pp. 53-65, 1999.

[8] Sánchez-Montañés, M. A., Verschure, P. M. F. J. and König, P., *Local and global gating of synaptic plasticity*, Neural Comp. 12 (3): 519-529, 2000.

[9] Körding, K. P. and König, P., *A single learning rule for the formation of efficient representations and for the segmentation of multiple stimuli*, Soc. Neurosci. Abst., 24, 323.16 (Abstract), 1998.